Uncanny...But Convincing? Inconsistency Between a Virtual Agent's Facial Proportions and Vocal Realism Reduces Its Credibility and Attractiveness, but Not Its Persuasive Success

Jan-Philipp Stein

Peter Ohler

Formal publication / citation:

Stein, J.-P., & Ohler, P. (2018). Uncanny...but convincing? Inconsistency between a virtual agent's facial proportions and vocal realism reduces its credibility and attractiveness, but not its persuasive success. *Interacting With Computers*. https://doi.org/10.1093/iwc/iwy023

Abstract

Embodied agents—i.e., digital systems represented by a virtual or robotic body—are used as persuasive tools in many different contexts. Still, psychological research indicates that for an agent to successfully influence its audience, many design factors have to work together to create a likable and trustworthy impression. Tapping into literature on the uncanny valley phenomenon, which has received only little attention in the field of persuasion research, we advance a consistency perspective that proposes matching levels of modality realism as a main requirement for users' acceptance. In an online experiment, we invite 107 participants to watch the persuasive speech of a virtual agent, manipulating both its facial proportions and vocal realism in a 2×2 between-subject design. Indeed, a mismatch between the realism of both features significantly reduces the agent's perceived credibility and attractiveness; yet, we observe that neither manipulation actually influences persuasive success in terms of attitude change. A potential explanation for this result pattern is offered by the Elaboration Likelihood Model, assuming that participants focused more on the agent's message than on peripheral cues to adjust their attitudes.

Keywords: virtual agent, persuasion, realism consistency, artificial voice, facial proportions, uncanny valley

Research Highlights

- Virtual agents are used as persuasive tools in various contexts (e.g., marketing).
- Uncanny valley research offers guidelines for well-accepted agent designs, highlighting the benefit of consistently realistic features.
- Realism consistency between facial and acoustic features improves subjective agent evaluations.
- However, actual attitude change is unrelated to feature consistency, probably relying more on message content.

Uncanny...But Convincing? Inconsistency Between a Virtual Agent's Facial Proportions and Vocal Realism Reduces Its Credibility and Attractiveness, but Not Its Persuasive Success

The success of persuasive efforts—i.e., those serving to influence another person's attitudes, beliefs, or behavior—rises and falls with various interpersonal perceptions. For instance, research has shown that attributions of authority (Cialdini, 2001), attractiveness (Davies, Goetz, & Shackelford, 2008), and social warmth (Gass & Seiter, 2015) all contribute in profound ways to the outcome of a persuasion attempt. Hence, a presumed expert who conveys a knowledgeable impression might be as convincing to an audience as a friendly speaker radiating with charisma—which is why both themes are highly popular in advertising contexts, fostering purchase intentions among potential customers.

However, empirical findings have demonstrated that the beneficial effects of perceived competence or likability not only apply to human-to-human interactions, but also hold true for persuasive messages conveyed by non-human entities (Fogg, 2003; Reeves & Nass, 1996). A much-noted subject in this regard are *embodied agents*: Digital systems that interact with human users through either a graphical (virtual agent) or a physical body (robot). Due to the immense technological advances of recent years, such anthropomorphic embodiments of technology now exist in the most diverse forms, ranging from simple 2D characters to highly complex 3D models and androids. Even more so, embodied agents have also diversified in terms of behavioral autonomy: While more basic versions remain confined to strict question-answer algorithms, others emerge as sophisticated artificial intelligences, employing machine-learning strategies to constantly increase their human likeness. Based on this variety of potential incarnations, interactive agents continue to be of great popularity in scenarios that require highly automated processes, but also rely on impressions of personal contact—including customer service, e-learning, or PR applications (Verhagen et al., 2014; Xu et al., 2014). In consequence, the trustworthiness and attractiveness of embodied agents has turned into a hot topic among psychology and HCI researchers, who strive to understand why the persuasion by a non-human entity is readily accepted in some cases—and strongly rejected in others.

The Role of Agent Appearance

Since visual attributes present a very obvious choice to distinguish between different types of agents, a particularly large number of studies have scrutinized the influence of agent appearance on users' willingness to accept a persuasive message. In a pioneering paper on collaborative virtual environments, Blascovich (2002) suggested that the social influence of a digital entity will likely increase with its visual realism; the more an agent resembles user expectations of a real human being, the higher may its persuasive power be. Although Blascovich's hypothesis has been supported by several empirical findings (e.g., Baylor, 2009; Guadagno et al., 2007; Roubroeks, Ham, & Midden, 2011), other research has raised doubts on its unlimited validity. For instance, a recent robotics experiment showed that robots with high degrees of human likeness were rated as less trustworthy than more mechanical-looking counterparts (Złotowski et al., 2015). Offering a possible interpretation, the authors locate their result within the famous "uncanny valley" framework (Mori, 1970), which suggests that increasing a technology's human likeness will only promote its likability up to a certain threshold before an 'eerie' ambiguity between machine and human may evoke strong unease among observers. Although many different explanations have been proposed to account for this phenomenon (for an introduction see MacDorman & Ishiguro, 2006), most uncanny valley studies emphasize the *eyes* of digital entities as a crucial impact factor on users' responses. For example, MacDorman and colleagues (2009) reported that atypical facial proportions with increased eye separation seemed highly disturbing to participants, thus reducing their acceptance of digital characters significantly. Similarly, researchers have discussed the detrimental effects of distorted eye size (Schindler et al., 2016; Seyama & Nagayama, 2007), eye vividness (Hanson, 2006; Tinwell, Abdel Nabi, & Charlton, 2013), and eye texture (MacDorman & Chattopadhyay, 2016). According to Schein and Gray (2015), observations like these ultimately relate to the fact that humans infer the most fundamental attributions (such as agency, emotional experience, or 'having a soul') from the eyes of an interaction partner; in consequence, the most refined 3D agent may turn out unexpectedly repulsive if the design of its eyes fails to match the anatomy of the human face, leading to an impression of mask-like inanimacy.

Surprisingly, an integration of the discussed uncanny valley findings into the literature on persuasive agents remains missing to this day. Instead, scholars of the latter have mostly focused on a more generalized understanding of *agent attractiveness*, which collects factors such as facial proportions (as well as many others) under one macro-level variable that is usually inferred from subjective ratings. The corresponding research indicates that artificial agents perceived as highly attractive might profit from the same beneficial effects that have been established for human-to-human contexts (e.g., Chaiken, 1979); for example, Holzwarth, Janiszewski, and Neumann (2006) reported that using an attractive virtual sales agent in the setting of a shopping website could greatly increase purchase intentions among potential customers. Along the same line, a study employing the hypothetical 'desert survival task' showed that an agent rated as highly attractive could induce much stronger attitude changes among participants than a moderately attractive alternative (Khan & Sutcliffe, 2014). According to Hanus and Fox (2015), the influence of such perceptions might be particularly strong if users are allowed to customize the virtual character with whom they are supposed to interact; similarly, other findings have suggested that an agent's persuasion attempt turns out most convincing if its physicality matches both the ethnicity (Pratt et al., 2007) and gender (Guadagno et al., 2007) of the human user.

While all of these findings certainly contribute to a better understanding of human-agent interaction, the current study argues that summarizing numerous visual aspects in a single container variable such as 'attractiveness' might conceal important effects on a lower level of abstraction. As such, we propose that studies focusing on more specific visual factors offer a welcome addition to the body of digital persuasion research.

The Role of Agent Voice and Feature Consistency

Even if developers expend their best efforts to design a visually pleasing persuasive agent, they will eventually notice that many other requirements have to be met to ensure their creation's success—after all, the agent's way of delivering its message adds many other variables to the equation. Since most applications of digital agents involve spoken dialogue, a growing body of research has explored the characteristics of *agent voices* as potential moderators of the audience's attitude change. For instance, a study using visually sexless robots uncovered that the perceived gender of an agent's voice could suffice to trigger corresponding stereotypes, thus supporting or impeding persuasive success (Powers & Kiesler, 2006). Examining an even broader main effect, other researchers have argued that merely adding a voice to an otherwise silent virtual agent could be enough to improve its perceived trustworthiness, even when using a simple text-to-speech voice (Qiu & Bensabat, 2004). This observation was corroborated in a later experiment by Stern, Mullennix, and Yaroslavsky (2006), who showed that increasing the human likeness of a computer's voice did not matter in terms of perceived credibility—its ability to speak was already sufficient to trigger the impression of social influence.

On the other hand, literature has also suggested that the artificiality of a virtual voice might turn out much more relevant if it is examined in interaction with different forms of visual presentation. Building towards a so-called *consistency effect*, a study conducted by Gong and Nass (2007) suggested that human voices may be rated as much more trustworthy when paired with a photorealistic face, whereas mechanical voices might be the preferable option in combination with a mechanomorphic appearance. Potential explanations for this phenomenon are numerous, especially if curious readers consult the decades' worth of consistency research that are offered by other areas of expertise. Evolutionary psychologists, for example, have provided striking evidence that stimulus consistency actually serves as a function of reproductive health and physical fitness (e.g., Klüver, Hecht, & Troje, 2015), while cognitive scientists connect the aversion against inconsistent entities to the undesirable neurological state of cognitive dissonance (e.g., Saygin et al., 2012). Again, further insight into the discussed effect is offered by uncanny valley researchers, whose work frequently indicates that digital entities with mismatching visual and auditory features tend to invite eeriness, fear, or disgust responses (e.g., MacDorman & Chattopadhyay, 2016; Mitchell et al., 2011; Tinwell, Grimshaw, & Abdel Nabi, 2015). Accordingly, Grimshaw (2009) even suggests that deliberate mismatches of modality quality offer unique design options to creators of horror media, who might find value in playing on the "primacy of the human voice" (p. 4). In any case, regardless of the theoretical approach that is used to explore inconsistency effects, the canonical conclusion remains that the human brain has a strong desire to perceive coherence in its surroundings—not least including the processing of persuasive messages from both human and human*like* communicators.

The Current Study

So far, most studies on agent-based persuasion have focused exclusively on over-arching trait attributions (i.e., attractiveness, trustworthiness) to uncover antecedents of participants' attitude change—which is then usually measured in a strictly hypothetical manner (e.g., "How would you behave when stranded in a desert?"). Also, while scholars of digital persuasiveness have compared humans to digital characters or robots to virtual agents, the role of subtle modifications within the same digital entity has received only little attention. As such, we identify three main shortcomings in previous studies on persuasive agents: (1) Potential effects of singular design aspects vanish as part of macro-level variables, (2) the obtained findings show limited ecological validity due to hypothetical-based measurements, and (3) design recommendations for specific agent features are usually impeded by a broad point of view. The current study was designed to overcome all of these research gaps. Not only did we include an actual screening of attitude change instead of 'what if'-type questions, we also acknowledged implications from the fascinating body of uncanny valley research, leading to the rather subtle manipulation of two selected agent characteristics; specifically, we focused on the proportions of a persuasive agent's eyes and the artificiality of its voice in order to examine consistency effects at a low level of abstraction.

In terms of our research hypotheses, we first addressed the most fundamental indicators of persuasive success: (a) the extent to which the target audience subjectively rates an argument as effective, and (b) the actual attitude change it evokes. Considering that realism consistency has been proposed as a strong influence on persuasiveness (Gong & Nass, 2007), we expected mismatches between the human likeness of an agent's facial proportions and its voice to result in significant damages to the two suggested criteria.

H1a: Mismatching the human likeness of a virtual agent's facial proportions and voice leads to a less favorable evaluation of its persuasive arguments.

H1b: Mismatching the human likeness of a virtual agent's facial proportions and voice decreases the attitude change caused by its persuasive arguments.

Secondly, we looked into various measures of social influence in regard to the virtual agent itself, as well as its way of delivering the persuasive message. For this purpose, we followed the theoretical groundwork by a pioneering study on persuasive agents (Stern, Mullennix, &

Yaroslavsky, 2006), which provided us with a measure for perceived *speaker credibility*. Unlike the variable explored by H1a, this construct focuses less on perceptions of the given arguments themselves, and more on attributions to the design and behavior of the persuading entity. Again, the scientific literature's negative outlook on entities with inconsistent features led us to hypothesize that mismatching the human likeness of our agent's eye region and voice should affect credibility evaluations in a negative way.

H2: Mismatching the human likeness of a virtual agent's facial proportions and voice decreases its perceived speaker credibility.

Lastly, we investigated two trait attributions that play an important role within the uncanny valley framework (Mori, 1970)—namely *eeriness* and *attractiveness*. Since inconsistency effects have been shown to strongly predict the uncanny aversion against humanlike entities (e.g., MacDorman & Chattopadhyay, 2016; Mitchell et al., 2011), we assumed that both ratings would turn out less favorable towards an entity with divergent facial and vocal human likeness.

H3a: Mismatching the human likeness of a virtual agent's facial proportions and voice increases its perceived eeriness.

H3b: Mismatching the human likeness of a virtual agent's facial proportions and voice decreases its perceived attractiveness.

Method

We devised an online experiment that revolved around a persuasive video message narrated by an embodied virtual agent. For a planned 2×2 between-subject design, we manipulated both the realism of the agent's voice (human voice vs. text-to-speech) and the design of its eye region (humanlike proportions vs. exaggerated eyes), resulting in four different video clips. With the help of a block randomization procedure, each participant of our online study was presented with one of the four video versions, before filling a set of questionnaires. In the following, we give details about all used measures, as well as every manipulation and exclusion that has occurred in our study.

Participants

To recruit a sufficient number of participants, we sent out invitation messages via various communication channels, including mailing lists at the local university, Facebook groups, and websites dedicated to the sharing of academic studies (all in German language). The necessary sample size was calculated a priori by means of G*Power software (assuming a moderate effect sizes of $f = 0.3$). Since our procedure included a video with sound, we made sure to instruct potential participants to prepare speakers or headphones before clicking on the provided link. A total of 128 participants responded to our call and completely filled in the online questionnaire. However, checking each dataset for the time that the persuasive video had been watched, we had to exclude 21 participants as their recorded values indicated a lack of attention by either falling below the video's duration or exceeding it by several minutes (with the specific thresholds chosen a priori).

According to our power analysis, the final sample of 107 participants was still suited to detect effects of moderate size with sufficient power. Exploring the recruited group in terms of demographics, we observed an age range of 18 to 54 years ($M = 24.8$ years, $SD = 5.45$), with gender distribution slightly skewed towards female participants (70.1% women, 29.9% men). Furthermore, regarding their educational level, most participants were either enrolled as students (42.1%) or had already acquired their university degree (50.5%). Even stronger homogeneity emerged in terms of religious background, as all participants filled in either atheism (59.8%) or Christian beliefs (40.2%) without exception.

Informed consent was obtained from every participant immediately at the start of the prepared online questionnaire. As a compensation for their time, participants were offered the chance to win one of two €25 gift cards, for which they had to enter their e-mail address in a separate questionnaire that could not be linked to their other responses.

Stimulus Materials

We created the basic model of our virtual agent with the graphics software *Adobe Fuse*, which allows for the creation of highly realistic digital characters. Adhering to the design choices of previous studies, we opted for an adolescent male with a friendly appearance (light clothing, subtle smile) and the software's preset parameters for humanlike facial dimensions. Since we were interested in comparing this natural version of our agent to a modification with unrealistic eye proportions, we subsequently modified the face of our initial model into a second version, using the highest values for iris scale, pupil size and eye distance available in Adobe Fuse (Fig. 1). To check the validity of the prepared manipulation, we conducted a short pretest, in which 32 independent student evaluators rated the human likeness of both models' facial proportions on a 7-point scale (1 = *not at all humanlike*, 7 = *extremely humanlike*). Having obtained a value of $M = 5.13$ for the realistic agent and $M = 2.41$ for the version with distorted facial proportions, a dependent t-test clearly indicated the significant dissimilarity of the prepared stimuli, $t(31) = 10.35$, $p < .01$, with a very large effect size of Cohen's $d = 1.83$.

Following the basic modeling process, we next researched an appropriate topic to be used in our agent's persuasive appeal. Above all else, we considered it important to choose a subject that was likely to avoid extreme initial attitudes among our participants in order to leave room for actual attitude change. Based on a review of several news topics as well as a focus talk with our colleagues, we eventually decided on the topic of "green genetic engineering" and scripted a four-minute speech surrounding this complex issue. Although the final manuscript acknowledged a few contra arguments (for the sake of authenticity), the bigger part of our text served to illustrate the "many advantages of green genetics technology if it is used in a regulated and thoughtful way"—therefore constituting a persuasive message in favor of the subject.

With our script finalized, we required both a humanlike as well as an artificial voice for the experimental manipulation of vocal realism. First, we hired a professional speech scientist and voice trainer to read our manuscript in a neutral, yet slightly friendly style and with the vocal pitch, timbre, and shimmer appropriate for a middle-aged male. Next, we entered the exact same text into the *Natural Reader* text-to-speech software, which offers a relatively sophisticated artificial voice. Despite the software's high quality, however, the resulting text-to-speech recording sounded undoubtedly mechanical due to its concatenated nature, therefore offering a fitting stimulus for the "artificial voice" condition. To avoid confounding effects from different stimulus durations, post-production was used to equalize the length of both recordings to a fixed length of 220 seconds. Additionally, a (very subtle) electronic echo was added to the artificial voice to make the manipulation even more salient.

Once the different stimulus components were fully prepared, we put all of them together into a four-minute animation of our virtual agent giving the persuasive speech, using the graphics engine *Unity* and the animation tool *Adobe Mixamo*. Apart from minimalistic head and hand movements that were added manually, most of the agent's movements were rendered by the plug-in *Salsa with Random Eyes 3D*, which automatically synchronizes a virtual character's lip movements to any given sound layer, while also adding natural blinking effects to the agent's eyes. By keeping this main animation constant—but swapping character models and sound files as needed—we were able to record four videos of the persuasive speech, corresponding to all required factor combinations: (a) proportional eyes/human voice, (b) proportional eyes/artificial voice, (c) disproportional eyes/human voice, and (d) disproportional eyes/artificial voice. Due to the applied step-by-step design process, all other variables of the video (e.g., background, duration of eye contact, body movements) remained standardized across conditions (Fig. 2).

Measures

Persuasive success. As a self-report measure of the persuasive speech's effectiveness, we used the six-item scale developed by Baker and Petty (1994), including the semantic differentials "harmful–beneficial", "unconvincing–convincing", "effective–ineffective", "negative–positive", "bad–good", and "foolish–wise". For sufficient differentiability, all items were presented in a seven-point answer format. The internal consistency of the resulting "argument quality" index turned out excellent (Cronbach's $\alpha = .94$).

To assess the actual attitude change evoked by the virtual agent's speech, we adopted a procedure suggested by Rosselli, Skelly, and Mackie (1995), which suggests comparing participants' opinions in identical pre- and post-stimulus questionnaires. For this purpose, we prepared three statements about the subject of our video, which had to be rated on 7-point scales (e.g., "Green genetics engineering may contribute to people's quality of life"; 1 = *fully disagree*, 7 = *fully agree*). Following the reverse-coding of one negatively worded statement, we averaged all three items into a composite attitude score. To eliminate the possibility that our stimuli or experimental setting might simply trigger an unspecific increase in agreeableness, we added three control topics with three items each (tuition fees, animal testing, industrial waste). In summary, our attitude assessment therefore consisted of four composite attitude scores (corresponding to four topics), measured once before and once after the persuasive video.

Qualities of the speaker. Participants' perceptions of the virtual agent and its way of speaking were measured using eight items (e.g., "evasive–straightforward", "insincere–sincere"; all rated on a 7-point scale) taken from Stern, Mullennix, and Yaroslavsky (2006). Although the authors originally suggest to use their items to measure two constructs—seven items for speaker credibility and one item for speaker competence—an exploratory factor analysis revealed that all items actually loaded very high on a single factor. In consequence, we combined them into a single speaker credibility index, which achieved an excellent Cronbach's α of .92. Yet another measure provided by the authors, a four-item scale on speaker strength (e.g. "timid–bold"), proved to be of unsatisfactory internal consistency, Cronbach's α = .48, and was therefore not included in this study.

Uncanny valley effect. Uncanny valley researchers have developed an abundance of possible operationalizations for the 'familiarity' effect initially proposed by Mori (1970). Among the most well-established measures are two indices developed by Ho and MacDorman (2010), who suggest that their scales are used in juncture to understand observers' affinity (or, from a negative perspective, aversion) towards a presented stimulus. Whereas the eeriness scale comprises eight items that relate to a weird, spine-tingling feeling (e.g., "reassuring–eerie"; "predictable–thrilling", Cronbach's α = .74), the attractiveness scale's five items refer more to aesthetic factors, including the semantic differentials "repulsive–agreeable" and "ugly–beautiful" (Cronbach's α = .90). Indeed, confirmatory factor analyses conducted with the data of our participants mostly supported the factorial structure suggested by Ho and MacDorman, indicating that the two measures actually addressed distinct latent variables with rather low covariance. However, we also found that one item of the eeriness scale hardly loaded on its respective factor; its subsequent exclusion resulted in a model that fit the data significantly better, $\chi^2(11) = 36.05$, $p < .01$, with the Comparative Fit Index (CFI) increasing from .90 to .94. As such, we ultimately used seven instead of eight items to compose the eeriness scale, improving internal consistency to Cronbach's α = .79 by this decision.

Manipulation checks and control variables. Although our experimental conditions were designed to be highly distinct from each other, we added manipulation checks to confirm the dissimilarity of the developed factor levels. For the facial proportions manipulation, we developed three semantic differentials that had to be answered in a seven-point format (e.g., "unnatural–natural eyes"), before averaging them into a single facial proportions score. Similarly, the realism of the agent's voice was measured with three self-developed items (e.g., "artificial–natural voice") that were combined into a vocal realism score. Both resulting indices proved to be of high to excellent internal consistency, Cronbach's $\alpha = .80$ and .90, respectively.

Lastly, to make sure that our between-subject design would not suffer from systematic disruptions in terms of participants' interest in—and comprehension of—our persuasive video message, we included two control variables, which were inspired by previous research (Stern, Mullennix, & Yaroslavsky, 2006). A two-item scale on message stimulation ("uninteresting–interesting", "boring–stimulating"; Cronbach's $\alpha = .73$) explored participants' interest in the presented message, while a single-item scale on message complexity ("simple–complex") addressed their understanding of the agent's monologue. Comparing the outcomes of these measures, we were able to establish that our four experimental groups were statistically equivalent in regard to their fascination for the topic of green genetic engineering and their understanding of the agent's speech.

Results

The statistics software *IBM SPSS 20* and the *lavaan* package in *R* were used to conduct all statistical investigations of this study. Providing a comprehensive overview of the collected data, Table 1 and Table 2 summarize the means and standard deviations for the obtained self-report measures, whereas Table 3 lists the results of the pre- and posttest attitude screening. Moreover, to allow for a direct comparison of the examined main and interaction effects, Table 4 collects the results from all conducted analyses of variance (ANOVAs).

Manipulation Checks and Control Variables

In order to check if our experimental manipulations had the desired effect on our participants' perceptions, we conducted separate independent t-tests for the two self-developed manipulation check scales. As expected, the t-test comparing the obtained facial proportions scores between participants who had watched the agent with realistic eyes (M = 4.34, SD = 1.29) and those who had seen the exaggerated version (M = 3.52, SD = 1.47) uncovered a significant group difference, $t(105) = 3.06$, $p < .01$, with a moderate effect size of Cohen's $d = 0.59$. Hence, we note that the agent's visual features indeed influenced participants' impression of humanlike facial proportions, supporting the validity of our manipulation.

Concerning the voice realism check, another independent t-test resulted in a very strong group difference between the experimental conditions with different sound design, $t(100.28) = 9.10$, $p < .01$, Cohen's $d = 1.77$. Indeed, participants who had heard the human speaker in the video expressed much higher vocal realism ratings (M = 4.38, SD = 1.42) than the groups listening to a text-to-speech voice (M = 2.09, SD = 1.16)—so that we also accept our second manipulation as successful.

To explore potential group differences in our additional control variables, *message stimulation* and *message* complexity, we entered them as dependent variables in two separate ANOVAs, using the facial and vocal realism conditions as between-subject factors. Doing so, neither analysis resulted in significant main or interaction effects. Although we cannot rule out the occurrence of type II errors in this regard, our power analysis suggests that the obtained sample size makes it rather unlikely to overlook effects, at least when assuming a medium effect size (87% power for $f = 0.3$). Hence, we argue that neither the perceived interestingness of the agent's speech, nor participants' comprehension of it seem to have differed strongly between the experimental groups, allowing us to disregard two potential confounding influences.

Persuasive Success

As a first criterion of persuasive success, we investigated whether the four experimental groups rated the quality of the persuasive arguments in a significantly different way. Conducting a two-way ANOVA with both experimental factors as independent variables, we examined no significant main effects, but a significant interaction between vocal realism and facial proportions, $F(1,103) = 4.33$, $p = .04$, $\eta_p^2 = .04$. Specifically, participants in the 'disproportional face/artificial voice' condition rated the presented arguments as almost perfectly effective ($M = 6.85$, $SD = 1.12$), followed by marginally lower ratings in the 'proportional face/human voice' condition ($M = 6.42$, $SD = 1.38$). At the same time, conditions with mismatching facial and vocal human likeness led to yet another drop in ratings, both for 'proportional face/artificial voice' ($M = 6.00$, $SD = 1.88$) and 'disproportional face/human voice' ($M = 5.97$, $SD = 1.89$). Although all obtained scores turned out rather high on the applied 7-point scale, we still report that realism inconsistency had a significant negative effect on the perceived quality of a persuasive argument as postulated by hypothesis H1a.

Secondly, we focused on participants' attitude changes as measured by our repeated approval screening. To make sure that the stimulus video had only affected attitudes about the relevant subject (green genetic engineering), we started our analysis by comparing the mean approval difference for this topic with the averaged difference across the three control topics (Tab. 3). Doing so, we observed that participants' opinion about the video subject had improved by $M = 0.71$ scale units (or approximately 10% of the 7-point scale), whereas their approval of the control topics only increased by $M = 0.06$ points (nearly 1% of the scale). Unsurprisingly, a dependent t-test marked this difference as statistically significant, $t(106) = 6.52$, $p < .01$. Having established the persuasive specificity of the presented video, we subsequently looked into the impact of both experimental factors on participants' attitude changes. However, the according two-way ANOVA did not result in any significant effects, neither a main effect of facial design,

$F(1,103) = 0.21$, $p = .65$, nor a main effect of voice design, $F(1,103) = 0.28$, $p = .60$, nor an interaction effect, $F(1,103) = 0.43$, $p = .52$. Instead, we report that participants showed similar changes in approval regardless of condition, ranging only slightly from $M = 0.64$ scale units ($SD = 1.09$) in the "proportional face/artificial voice" condition to $M = 0.87$ scale units ($SD = 1.10$) in the "proportional face/human voice" condition. In light of this, we reject hypothesis H1b: Inconsistency between facial proportions and vocal realism was not shown to influence actual attitude change in a statistically significant way.

Speaker Credibility

Addressing our next hypothesis, we conducted another univariate ANOVA, focusing on the developed index for *speaker credibility* as the dependent variable. The procedure resulted in no significant main effects of facial and voice design, but uncovered a significant interaction, $F(1,103) = 5.39$, $p = .02$, $\eta_p^2 = .05$. As hypothesized, an inconsistency between both manipulated attributes resulted in notably lower credibility attributions, with almost the same mean ratings found in the 'proportional face/artificial voice' ($M = 4.88$, $SD = 1.11$) as in the 'disproportional face/human voice' condition ($M = 4.88$, $SD = 1.12$). In contrast to this, the two consistent conditions yielded slightly higher scores, regardless of whether the manipulated attributes were presented as coherently humanlike ($M = 5.28$, $SD = 0.75$) or artificial ($M = 5.38$, $SD = 0.92$). As such, hypothesis H2 is supported by our data, indicating a clear benefit of feature consistency for subjective credibility perceptions.

Uncanny Valley Effect

Concluding our statistical analyses, we entered the final pair of measures—the uncanny valley concepts *eeriness* and *attractiveness*—into two final ANOVAs. To control for family-wise error, we adjusted the threshold of significance using Bonferroni correction, so that only values below $p < .025$ were regarded as statistically significant for these measures.

For eeriness, the respective ANOVA revealed no significant effects: No main effect for facial design, $F(1,103) = 0.94$, $p = .34$, or voice design, $F(1,103) = 0.61$, $p = .44$, as well as no significant interaction, $F(1,103) = 0.17$, $p = .69$. On the other hand, our exploration of the obtained attractiveness ratings—while also missing significant main effects—yielded a highly significant interaction effect between facial and vocal design, $F(1,103) = 11.95$, $p < .01$, with a large effect size of $\eta_p^2 = .10$. A follow-up examination of the specific group means revealed that consistent conditions were both perceived as highly attractive, with $M = 5.02$ ($SD = 0.83$) for the 'proportional face/human voice' condition, and $M = 4.94$ ($SD = 1.01$) for the condition with matching features of increased artificiality. In contrast to this, inconsistencies in the agent's design led to notable decreases in perceived attractiveness, both for the 'proportional face/artificial voice' ($M = 4.28$, $SD = 1.45$) and the 'disproportional face/human voice' variants ($M = 4.25$, $SD = 0.87$). In summary, we report evidence in favor of hypothesis H3b, but negative conclusions regarding H3a: Whereas attractiveness ratings clearly suffered from inconsistency effects, eeriness perceptions remained mostly unrelated to the variations of our agent's face and voice.

Intercorrelations Between Outcome Variables

For a better understanding of our obtained data pattern, Table 5 presents the zero-order correlations between all dependent variables. Focusing on the strongest coefficients uncovered by our calculations, we note that attractiveness was associated rather strongly with speaker credibility ($r = .61$). Speaker credibility, in turn, was found to be clearly related to perceptions of argument quality ($r = .67$), indicating that the mere trustworthiness of an entity may suffice to make the content of its arguments seem convincing and effective—or vice versa. Of course, the correlational nature of the findings presented in Table 5 asks for a cautious interpretation of the direction of these relationships.

Discussion

Persuasive non-human entities are not exclusively a present-day phenomenon—after all, puppets and similar creations have been used for centuries to nudge children towards new ideas or desired behaviors. Still, as digital technology continues to branch out into the most diverse application fields, artificial 'influencers' have become a subject more relevant than ever, for developers and researchers alike. So far, both theoretical and empirical endeavors trying to shed light on digital persuasiveness have reached the same conclusion: Not only do interactions with embodied agents depend on how they look or sound, but every single action and audience characteristic adds another confound to the mix. Despite this confusing web of variables, however, scholars have still managed to flesh out some basic principles that improve the acceptance of persuasive agents—including the human desire to perceive consistency in virtual humans (Gong & Nass, 2007), or the particular importance surrounding digital eyes (Schein & Gray, 2015).

Varying both the facial proportions and vocal realism of a virtual agent in the context of a persuasive video on green genetics, we found consistency effects for three out of five hypotheses. As expected, the perceived argument quality, speaker credibility, and agent attractiveness all increased if a coherent design was chosen—that is, if a proportional face was paired with a human voice, or a text-to-speech voice was added to a more exaggerated facial model. To our surprise, however, the same did not apply to perceived eeriness or the resulting attitude change, both of which could not be linked statistically to our manipulation of feature consistency.

Examining the intercorrelations between our outcome variables helps to make some sense of the observed results. The high correlation coefficients between speaker attractiveness, speaker credibility, and argument quality are easy to locate within previously established norms of successful persuasion (e.g., Patzer, 1983), which frame these perceptions as intrinsically linked.

At the same time, our correlational insight does not provide a clear answer as to why participants' attitude change remained unrelated to our experimental manipulations. In our interpretation, a better explanation for this particular result is offered by the Elaboration Likelihood Model (ELM; Petty, & Cacioppo, 1984).

Providing a well-established social psychological approach to phenomena of social influence, the ELM argues that persuasive success relies on both a central and a peripheral route of information processing. Whereas the former path is paved by thoughtful considerations of the persuasive message's logic and rational merit, the latter gets shaped by general perceptions of the message's source (e.g., the attractiveness of its sender) and style (e.g., message simplicity). Additionally, the model suggests that people mostly use the central route as long as they have enough time and motivation to process the given information. For our experiment, we have to assume that this was the case: Not only did we ask participants to watch the video attentively, but the simple fact that they were to expect a post-video questionnaire must also have contributed to a stronger engagement with the presented message. As such, we suppose that the peripheral route—thus, the factors that were actually varied in this study—was likely pushed to the background, accounting for the lack of difference between the experimental groups' attitude change.

Adding to this main explanation, we would also like to highlight the vast number of confounding variables that eventually contribute to the adoption, modification, and dismissal of attitudes, such as prior knowledge or certain social expectations. For instance, a person might find a speaker attractive or credible, but still possess conflicting knowledge or have other interpersonal preferences that override the persuasive effects of these traits. Previous research has further suggested that users' personality (von der Pütten, Krämer, & Gratch, 2010), their level of social ostracism (Ruijten, Midden, & Ham, 2015), and the expectation of specific social cues (Ruijten, Midden, & Ham, 2016) all strongly influence the way virtual characters and their persuasion attempts are perceived. In consequence, we find it safe to assume that our participants' inclination to accept the agent's message depended on many more factors than those measured in our study. For future studies, we therefore suggest additional efforts to assess participants' initial knowledge about a topic, as well as some form of personality and mood assessment to control for more confounding effects.

On a more conclusive note, we argue that our findings provide fascinating tie-ins for researchers interested in the uncanny valley phenomenon. Contrary to our initial idea, we found that only our agent's attractiveness—but not its eeriness—was subject to a clear consistency effect. Whereas the observed attractiveness result clearly concurs with previous studies from the field of evolutionary psychology (e.g., Klüver, Hecht, & Troje, 2015), the lacking connection between feature consistency and eeriness raises some conceptual questions, especially since it directly opposes the results of recent experiments (MacDorman & Chattopadhyay, 2016; Mitchell et al., 2011; Tinwell, Grimshaw, & Abdel Nabi, 2015). Even more so, examining the zero-order correlations between our variables, we come to the surprising conclusion that eeriness was actually positively related to facial and vocal realism indices, as well as perceived attractiveness. In all probability, these paradoxical findings indicate that the established conceptualization of eeriness might not be entirely feasible to produce stable effects across different types of stimuli and audiences—a criticism that is echoed by previous methodological reviews (Kätsyri et al., 2015). Moreover, we would like to point out that our specific scenario was obviously less prone to eerie feelings than anticipated, which is underscored by low ratings across all four conditions. A potential cause for this, at least in our understanding, might lie in the abilities of the designed virtual agent. According to recent research, the acceptance of non-human entities is actually strongly influenced by their mental abilities (Gray & Wegner, 2012) and autonomy (Złotowski, Yogeeswaran, & Bartneck, 2017). As such, artificially intelligent systems might evoke much more aversive reactions than scripted agents such as the one used in this study, whose predictable behavior should appear significantly less threatening—i.e., less uncanny—to observers (Stein & Ohler, 2017). Taken together with the fact that the 'disproportional face/artificial voice' condition yielded the most favorable ratings in half of our dependent variables, we suggest that a highly artificial *scripted* agent might even trigger explicitly beneficial schemata, such as the expectation of precision and infallibility typically evoked by computer systems. In any case, we report that eeriness perceptions could not be connected to any other outcome variable in our study, neither to speaker credibility or argument quality, nor to observable attitude changes. At last, this might mean good news to designers of persuasive agents; if our observation holds true in other contexts, developers have to be less afraid of their creations falling into uncanny valley territory, as long as some basic form of attractiveness remains. If our results are taken into consideration, consistent designs are probably a good thing to aim for in this regard.

Limitations

Despite our best efforts to recruit an adequate sample and to compose sophisticated stimulus materials, further replications and theoretical modifications will surely help to come to a more comprehensive understanding of the factors surrounding virtual agents, their persuasiveness, and possible uncanny valley effects. In particular, additional investigations with samples from different cultural backgrounds might be needed to account for the strong cross-cultural differences in technology acceptance that have been reported by literature (Kaplan, 2004; Rau, Li, & Li, 2009). Since nearly all participants reported a high level of education and either an atheistic or Christian background, we suppose that our results cannot be generalized for the whole sociocultural spectrum. Also, investigations of participants' technical expertise could be added to further research endeavors to control for potential habituation or novelty effects. Most of all, however, we acknowledge that attitude changes cannot necessarily be fostered by one or two modality manipulations, so that any future research on persuasive agents will be of great value—not only to academics, but also to decision makers in customer service, advertising, and education contexts.

References

Baker, S. M., & Petty, R. E. (1994). Majority and minority influence: Source position imbalance as a determinant of message scrutiny. *Journal of Personality and Social Psychology, 67*, 5–19.

Baylor, A. L. (2009). Promoting motivation with virtual agents and avatars: Role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1535), 3559–3565.

Blascovich, J. (2002). A theoretical model of social influence for increasing the utility of collaborative virtual environments. In *Proceedings of the 4th International Conference on Collaborative Virtual Environments* (pp. 25–30). Bonn, Germany.

Chaiken, S. (1979). Communicator physical attractiveness and persuasion. *Journal of Personality and Social Psychology*, *37*(8), 1387–1397.

Cialdini, R. B. (2001). *Influence: Science and practice*. Boston: Allyn & Bacon.

Davies, A. P. C., Goetz, A. T., & Shackelford, T. K. (2008). Exploiting the beauty in the eye of the beholder: The use of physical attractiveness as a persuasive tactic. *Personality and Individual Differences*, *45*, 302–306.

Fogg, B. J. (2003). *Persuasive technology*. Burlington: Morgan Kaufmann.

Gass, R. H., & Seiter, J. S. (2015). *Persuasion: Social influence and compliance gaining*. Abingdon: Routledge.

Gong, L., & Nass, C. (2007). When a talking-face computer agent is half-human and half-humanoid: Human identity and consistency preference. *Human Communication Research, 33*, 163–193.

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*, 125–130.

Grimshaw, M. (2009). The audio Uncanny Valley: Sound, fear and the horror game. In *Proceedings of the 4th Conference on Interaction with Sound, Audio Mostly 2009* (pp. 21–26). Glasgow, UK.

Guadagno, R. E., Blascovich, J., Bailenson, J. N., & Mccall, C. (2007). Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology*, *10*(1), 1–22.

Hanson, D. (2006). Exploring the aesthetic range for humanoid robots. In *Proceedings of the ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*, Vancouver, Canada (pp. 16–20).

Hanus, M. D., & Fox, J. (2015). Persuasive avatars: The effects of customizing a virtual saleperson's appearance on brand liking and purchase intentions. *International Journal of Human-Computer Studies, 84*, 33–40. doi: 10.1016/j.ijhcs.2015.07.004

Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, *26*, 1508–1518.

Holzwarth, M., Janiszewski, C., & Neumann, M. (2006). The influence of avatars on online consumer shopping behavior. *Journal of Marketing*, *70*, 19–36. doi: 10.1509/jmkg.70.4.19

Kaplan, F. (2004). Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. *International Journal of Humanoid Robotics, 1(3)*, 465–480.

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, *6*, 390. doi: 10.3389/fpsyg.2015.00390

Khan, R. F., & Sutcliffe, A. (2014). Attractive agents are more persuasive. *International Journal of Human-Computer Interaction, 30*(2), 142–150.

Klüver, M., Hecht, H., & Troje, N. F. (2015). Internal consistency predicts attractiveness in biological motion walkers. *Evolution and Human Behavior, 37*(1), 40–46.

MacDorman, K. F., & Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition, 146*, 190–205.

MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior, 25,* 695–710.

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies, 7*(3), 297–337. doi:10.1075/is.7.3.03mac

Mitchell, W. J., Szerszen, K. A., Sr., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception, 2,* 10–12. doi: 10.1068/i0415

Mori, M. (1970). Uncanny valley. *Energy*, *7*(4), 33–35.

Patzer, G. L. (1983). Source credibility as a function of communicator physical attractiveness. *Journal of Business Research*, *11*(2), 229–241.

Petty, R. E., & Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, *46*, 69–81.

Powers, A., & Kiesler, S. (2006). The advisor robot: Tracing people's mental model from a robot's physical attributes. In *Proceedings of the ACM/IEEE International Conference on Human–Robot Interaction* (pp. 218–220), Salt Lake City, USA.

Pratt, J. A., Hauser, K., Ugray, Z., & Patterson, O. (2007). Looking at human-computer interface design: Effects of ethnicity in computer agents. *Interacting with Computers, 19*(4), 512–523.

Qiu, L., & Bensabat, I. (2005). The effects of text-to-speech voice and 3D avatars on consumer trust in the design of live help interface of electronic commerce. In *Proceedings of the 10th Americas Conference on Information Systems* (pp. 3165–3173). New York, USA.

Rau, P. L. P., Li, Y., & Li, D. (2009). Effects of communication style and culture on ability to accept recommendations from robots. *Computers in Human Behavior, 25,* 587–595.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge: Cambridge University Press.

Rosselli, F., Skelly, J. J., & Mackie, D. M. (1995). Processing rational and emotional messages: the cognitive and affective mediation of persuasion. *Journal of Experimental Social Psychology*, *31*, 163–190.

Roubroeks, M., Ham, J., & Midden, C. J. H. (2011). When artificial social agents try to persuade people: The role of social agency on the occurence of psychological reactance. *International Journal of Social Robotics, 3,* 155–165.

Ruijten, P. A. M., Midden, C. J. H., & Ham, J. (2015). Lonely and susceptible: The influence of social exclusion and gender on persuasion by an artificial agent. *International Journal of Human-Computer Interaction, 31*(11), 832–842. doi:10.1080/10447318.2015.1067480

Ruijten, P. A. M., Midden, C. J. H., & Ham, J. (2016). Ambiguous agents: The influence of consistency of an artificial agent's social cues on emotion recognition, recall, and persuasiveness. *International Journal of Human-Computer Interaction, 32*(9), 734–744. doi:10.1080/10447318.2016.1193350

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience, 7(4)*, 413–422.

Schein, C., & Gray, K. (2015). The eyes are the window to the uncanny valley. Mind perception, autism and missing souls. *Interaction Studies, 16*(2), 173–179.

Schindler, S., Zell, E., Botsch, M., & Kissler, J. (2017). Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. *Scientific Reports, 7,* 45003. doi:10.1038/srep45003

Seyama, J., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. Presence, 16(4), 337–351. doi:10.1162/pres.16.4.337

Stein, J.-P., & Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition, 160*, 43–50. doi: 10.1016/j.cognition.2016.12.010

Stern, S. E., Mullennix, J. W., & Yaroslavsky, I. (2006). Persuasion and social perception of human vs. synthetic voice across person as source and computer as source conditions. *International Journal of Human-Computer Studies, 64*, 43–52.

Tinwell, A., Abdel Nabi, D., & Charlton, J. (2013). Perception of psychopathy and the uncanny valley in virtual characters. Computers in Human Behavior, 29, 1617–1625. doi:10.1016/j.chb.2013.01.008

Tinwell, A., Grimshaw, M., & Abdel Nabi, D. (2015). The effect of onset asynchrony in audio-visual speech and the Uncanny Valley in virtual characters. *International Journal of Mechanisms and Robotic Systems*, *2*(2), 97–110.

Verhagen, T., van Nes, J., Feldberg, F., & van Dolen, W. (2014). Virtual customer service agents: Using social presence and personalization to shape online service encounters. *Journal of Computer-Mediated Communication, 19*(3), 529–545.

von der Pütten, A., Krämer, N. C., & Gratch, J. (2010). How our personalities shape our interactions with virtual characters. Implications for research and development. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, & A. Safonova (eds.), *Intelligent Virtual Agents* (pp. 208–221). Berlin: Springer.

Xu, D., Huang, W. W., Wang, H., & Heales, J. (2014). Enhancing e-learning effectiveness using an intelligent agent-supported personalized virtual learning environment: An empirical investigation. *Information & Management, 51*(4), 430–440.

Złotowski, J., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2015). Persistence of the uncanny valley: The influence of repeated interactions and a robot's attitude on its perception. *Frontiers in Psychology*, *6*, 883. doi:10.3389/fpsyg.2015.00883

Złotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies*, *100,* 48–54. doi:j.ijhcs.2016.12.008

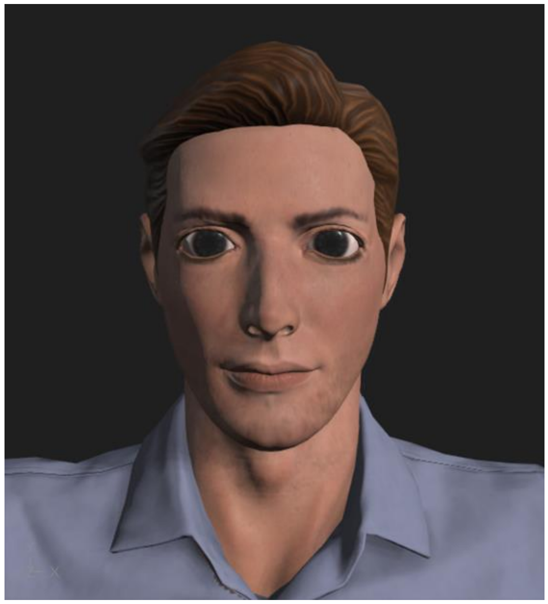

Figure 1. Final design of the study’s virtual agent with exaggerated facial proportions (focusing specifically on the eye region).

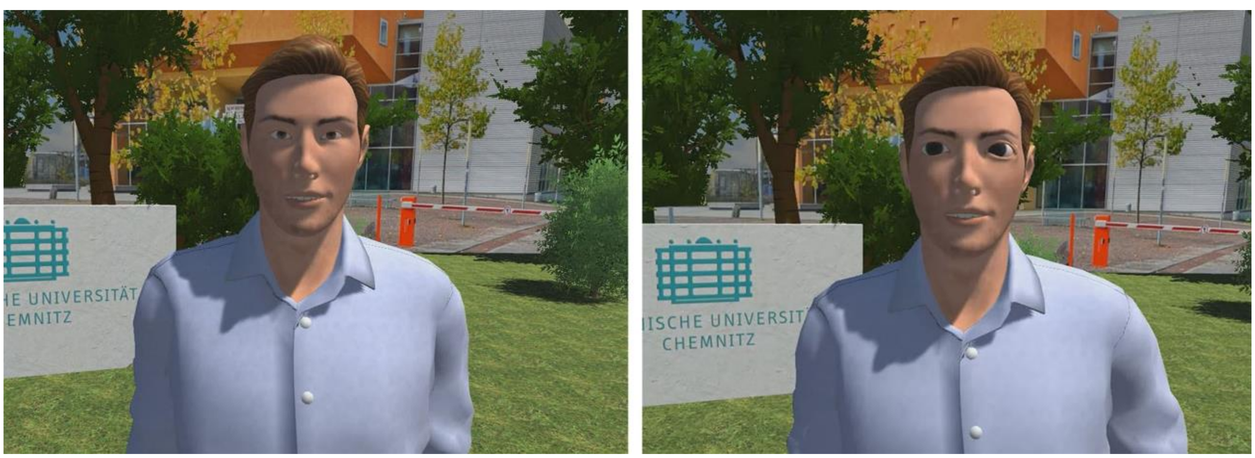

Figure 2. A comparison between the persuasive agent with humanlike facial proportions (left) and the agent with exaggerated eyes (right). All other visual aspects of the recorded scene are standardized across conditions.

Table 1. *Means and standard deviations for the manipulation check scales and two control variables*

	proportional face				disproportional face			
	realistic voice		artificial voice		realistic voice		artificial voice	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
facial proportions scale	4.48	1.19	4.21	1.38	3.14	1.40	3.91	1.45
vocal realism scale	4.79	1.11	1.94	1.16	4.02	1.59	2.25	1.17
message stimulation	4.36	1.45	4.48	1.16	4.14	1.24	4.67	1.49
message complexity	3.60	1.22	3.63	1.73	3.18	1.66	3.29	1.51

Note. All indices range from 1 to 7.

Table 2. *Means and standard deviations for the study's self-report measures.*

	proportional face				disproportional face			
	realistic voice		artificial voice		realistic voice		artificial voice	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
quality of argument	6.42	1.38	6.00	1.88	5.97	1.89	6.85	1.12
speaker credibility	5.28	0.75	4.88	1.11	4.88	1.12	5.38	0.92
eeriness	2.89	0.62	2.70	0.74	2.97	1.04	2.95	0.77
attractiveness	5.02	0.83	4.28	1.45	4.25	0.87	4.94	1.01

Note. All indices range from 1 to 7.

Table 3. *Pre- and posttest approval for the topic of the persuasive video (genetic engineering) as well as three control topics*

	proportional face						disproportional face						all groups
	realistic voice			artificial voice			realistic voice			artificial voice			
	pre	**post**	**Δ**	**pre**	**post**	**Δ**	**pre**	**post**	**Δ**	**pre**	**post**	**Δ**	**Δ**
genetic engineering	3.79	4.65	**0.86**	3.91	4.55	**0.64**	3.81	4.46	**0.65**	3.88	4.56	**0.68**	**0.71**
control topics													
tuition fees	3.13	3.40	**0.27**	3.33	3.54	**0.21**	3.30	3.35	**0.05**	3.59	3.60	**0.01**	**0.13**
animal testing	4.03	4.16	**0.13**	4.52	4.51	**-0.01**	4.11	4.23	**0.12**	3.28	3.43	**0.15**	**0.10**
environment protection	5.01	5.00	**-0.01**	4.79	4.71	**-0.07**	4.88	4.92	**0.04**	5.07	5.01	**-0.06**	**-0.03**
average	4.06	4.19	**0.13**	4.21	4.25	**0.04**	4.10	4.17	**0.07**	3.98	4.01	**0.03**	**0.06**

Note. All scales range from 1 to 7.

Table 4. *Effects calculated in two-way ANOVAs focusing on the different outcome variables*

	main effect of face		**main effect of voice**		**interaction effect face × voice**	
	F	*p*	*F*	*p*	*F*	*p*
quality of argument	0.41	.53	0.54	.47	4.33	**.04***
attitude change	0.21	.65	0.28	.60	0.43	.52
speaker credibility	0.07	.80	0.07	.80	5.39	**.02***
eeriness	0.94	.34	0.61	.44	0.17	.69
attractiveness	0.08	.78	0.02	.90	11.95	**.00****

Notes. * $p < .05$, ** $p < .01$.

Table 5. *Zero-order correlations between measured variables*

	1	**2**	**3**	**4**	**5**	**6**	**7**
1 facial proportions scale	–						
2 vocal realism scale	.23	–					
3 argument quality	**.42***	.19	–				
4 attitude change	**.31***	.15	**.46***	–			
5 speaker credibility	**.45***	**.28***	**.67***	**.36***	–		
6 eeriness	**.27***	**.27***	.09	.00	.13	–	
7 attractiveness	**.66***	**.31***	**.48***	.22	**.61***	**.34***	–

Notes. * p < .001 (adjusted for multiple testing)